

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:16:50 ; Search time 91.75 Seconds

(without alignments)  
16.543 Million cell updates/sec

Title: US-09-331-631a-7\_COPY\_34\_80

Perfect score: 258  
Sequence: 1 TERDPROQYEQOCRCRCESEA.....QCEORCEREYEQORQDEEE 47

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	124	48.1	588	1	VCLB_GOSHI
2	124	48.1	605	1	VCLB_GOSHI
3	87	33.7	429	1	APPA_MACFA
4	86	33.3	47	1	AGRP_LUCFY
5	86	33.3	1898	1	TRHY_HUMAN
6	84.5	32.8	1407	1	TRHY_RABIT
7	79.5	30.8	401	1	APPA_PAPAN
8	79	30.6	743	1	ABRA_PAPAN
9	79	30.6	1905	1	TACR_DICDI
10	78	30.2	892	1	IF2_SALTY
11	78	30.2	905	1	SNF5_YEAST
12	78	30.2	1403	1	PRO_DROME
13	77.5	30.0	502	1	SKS1_YEAST
14	77	29.8	339	1	TF2D_HUMAN
15	77	29.8	857	1	NFM_CHICK
16	76	29.5	482	1	U2R2_HUMAN
17	76	29.5	648	1	KAPC_DICDI
18	76	29.5	890	1	IF2_ECOLI
19	75.5	29.3	1023	1	GLT_DROME
20	75	29.1	241	1	YOR8_CAEFL
21	75	29.1	758	1	YH38_YEAST
22	74.5	28.9	1549	1	TRHY_SHEEP
23	74	28.7	877	1	INCE_CHICK
24	73.5	28.5	1344	1	IF3A_MOUSE
25	73.5	28.5	1382	1	IF3A_HUMAN
26	73	28.3	418	1	YSE2_CAEFL
27	73	28.3	678	1	GARP_PLATF
28	73	28.3	1023	1	CLOC_DROME
29	72.5	28.1	287	1	TRT2_HUMAN
30	72	27.9	407	1	IE68_HSVSA
31	71.5	27.7	462	1	U2R2_MOUSE
32	71.5	27.7	1403	1	YDR3_SCHPO
33	71	27.5	510	1	CF2_DROME

34	71	27.5	514	1	CF23_DROME
35	71	27.5	966	1	SSN6_YEAST
36	71	27.5	2124	1	Y192_HUMAN
37	70.5	27.3	255	1	LP61_EIMTE
38	70.5	27.3	503	1	LEG1_PEA
39	70.5	27.3	1391	1	MSW2_DROHY
40	70	27.1	917	1	IF2_PROVU
41	69.5	26.9	298	1	TRT2_PROVU
42	69.5	26.9	300	1	TRT2_MOUSE
43	69.5	26.9	466	1	CYP8_CAEFL
44	69.5	26.9	1240	1	YNI1_YEAST
45	69	26.7	338	1	LEB8_PEA

## ALIGNMENTS

RESULT 1  
VCLB\_GOSHI STANDARD; PRT; 588 AA.  
AC P09801;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE VICILIN C72 PRECURSOR (ALPHA-GLOBULIN B).  
OS Gossypium hirsutum (Upland cotton).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids II; Malvales; Malvaceae; Gossypium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Chlan C.A., Pyle J.B., Legocki A.B., Dure L. III;  
RT "Developmental biochemistry of cottonseed embryogenesis and germination. XVII. cDNA and amino acid sequences of the members of the storage protein families".  
RT Plant Mol. Biol. 7:475-489(1986).  
CC -I- FUNCTION: SEED STORAGE PROTEIN.  
CC -I- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN BODIES.  
CC -I- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN, CONVICILIN, CONGLYCININ, ETC.).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL: M16891; AAA33071.1; -.  
DR PIR: A30838; FMCNAB.  
DR HSSP: P50477; ICANB.  
DR INTERPRO: IPR001113; -.  
DR PFAM: PF00546; Seedstore\_7s; 1.  
FT SIGNAL 1 25  
FT CHAIN 26 588 VICILIN C72.  
SQ SEQUENCE 588 AA: 69729 MW: 63B699B29AB8ADEB CRC64:

Query Match 48.1%; Score 124; DB 1; Length 588;  
Best Local Similarity 46.3%; Pred. No. 7.6e-05;  
Matches 19; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

OY 4 DPROQYEQOCRCRCESEATEEREOQCEORCEREYEQORQ 44  
DB 36 DPKRYEDCRRCRCESEATEEREOQCEORCEREYEQORQ 76

RESULT 2  
VCLB\_GOSHI STANDARD; PRT; 605 AA.  
ID VCLB\_GOSHI

```

AC P09799;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE VICILIN GC72-A PRECURSOR (ALPHA-GLOBULIN A).
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
RN [1]
RP SEQUENCE FROM N.A.
RA Chlan C.A., Borrieto K., Kamalay J.A., Dure L. III;
RT "Developmental biochemistry of cottonseed embryogenesis and
RT germination. XIX. Sequences and genomic organization of the alpha
RT globulin (vicilin) genes of cottonseed."
RL Plant Mol. Biol. 9:533-546(1987).
CC -1- FUNCTION: SEED STORAGE PROTEIN.
CC -1- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC BODIES.
CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICILIN, CONGLYCININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL: M19378; AAA33069.1; -.
CC PIR: S06398; S06398.
CC DR HSSP: P50477; ICAX.
CC DR INTERPRO: IPR001113; -.
CC DR PFAM: PF00546; Seedstore_7s; 1.
CC KW Seed storage protein; Signal.
CC SIGNAL 1 23
CC CHAIN 24 605 VICILIN GC72-A.
CC FT SEQUENCE 605 AA; 71049 MW; C9DB9371C976953B CRC64;
SQ

Query Match 48.1%; Score 124; DB 1: Length 605;
Best Local Similarity 43.5%; Pred. No. 7.8e-05;
Matches 20; Conservative 13; Mismatches 13; Indels 0; Gaps 0;

QY 2 EBDPQOYECORCESEATEEREOECORCEREYKEQOOROE 47
DB 34 EDDPQOYEDCKRKRCOLETRGQTEDEKCDRESETOLEKEQORDGED 79

RESULT 3
AP04_MACFA STANDARD; PRT: 429 AA.
ID AP04_MACFA
AC P33621;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE APOLIPOPROTEIN A-IV PRECURSOR (APO-AIV).
GN APOA4.
OS Macaca fascicularis (Crah eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=LEUKOCYTE;
RC MEDLINE=93192330; PubMed=8448212;
RA Osada Y., Pocovi M., Nicolosi R.J., Schaefer E.J., Ordovas J.M.;
RT "Nucleotide sequences of the Macaca fascicularis apolipoprotein C-II
RT and A-IV genes."
RL Biochim. Biophys. Acta 1172:335-339(1993).
CC -1- FUNCTION: MAY HAVE A ROLE IN CHYLOMICRONS AND VLDL SECRETION AND
CC CATABOLISM. REQUIRED FOR EFFICIENT ACTIVATION OF LIPOPROTEIN

```

```

CC LIPASE BY APOC-II; POTENT ACTIVATOR OF LCAT. APOA-IV IS A MAJOR
CC COMPONENT OF HDL AND CHYLOMICRONS.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -1- TISSUE SPECIFICITY: SECRETED IN PLASMA.
CC -1- DOMAIN: NINE OF THE THIRTEEN 22-AMINO ACID TANDEM REPEATS (EACH
CC 22-MER IS ACTUALLY A TANDEM ARRAY OF TWO, A AND B, RELATED II-
CC MERS) OCCURRING IN THIS SEQUENCE ARE PREDICTED TO BE HIGHLY ALPHA-
CC HELICAL, AND MANY OF THESE HELICES ARE AMPHIPATHIC. THEY MAY
CC THEREFORE SERVE AS LIPID-BINDING DOMAINS WITH LECITHIN:CHOLESTEROL
CC ACYLTRANSFERASE (LCAT) ACTIVATING ABILITIES.
CC -1- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL: X68361; CAA48421.1; -.
CC PIR: S29565; S29565.
CC DR PIR: S30195; S30195.
CC DR HSSP: P02649; INFO.
CC DR INTERPRO: IPR00074; -.
CC DR PFAM: PF01442; Apolipoprotein; 1.
CC KW Plasma; Lipid transport; HDL; Chylomicron; Repeat; Signal.
CC SIGNAL 1 20
CC CHAIN 21 429 BY SIMILARITY.
CC FT DOMAIN 33 330 13 X 22 AA APPROXIMATE TANDEM REPEATS.
CC FT REPEAT 33 54 1.
CC FT REPEAT 60 81 2.
CC FT REPEAT 82 103 3.
CC FT REPEAT 115 136 4.
CC FT REPEAT 137 158 5.
CC FT REPEAT 159 180 6.
CC FT REPEAT 181 202 7.
CC FT REPEAT 203 224 8.
CC FT REPEAT 225 246 9.
CC FT REPEAT 247 268 10.
CC FT REPEAT 269 286 11.
CC FT REPEAT 287 308 12.
CC FT REPEAT 309 330 13.
CC FT DOMAIN 372 420 GLU/GLN-RICH.
CC FT SEQUENCE 429 AA; 49876 MW; 3D458F51D0DB60C CRC64;
SQ

Query Match 33.7%; Score 87; DB 1: Length 429;
Best Local Similarity 34.8%; Pred. No. 0.13;
Matches 16; Conservative 16; Mismatches 14; Indels 0; Gaps 0;

QY 2 EBDPQOYECORCESEATEEREOECORCEREYKEQOOROE 47
DB 374 EQRREQQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQ 419

RESULT 4
ACRP_LUFVCY STANDARD; PRT: 47 AA.
ID AGRP_LUFVCY
AC P56568;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 6.5 KDA ARGININE/GLUTAMATE-RICH POLYPEPTIDE (6.5K-AGRP).
OS Luffa cylindrica (Smooth loofah) (Sponge gourd).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Cucurbitales; Cucurbitaceae; Luffa.
RN [1]
RP SEQUENCE.
RC TISSUE=SEED;
RA MEDLINE=97357433; PubMed=9214759;
RX Kimura M., Park S.S., Sakai R., Yamasaki N., Funatsu G.;

```

RT "Primary structure of 6.5k-arginine/glutamate-rich polypeptide from  
 the seeds of sponge gourd (Luffa cylindrica).";  
 RL Biosci. Biotechnol. Biochem. 61:984-988(1997).  
 CC -1- FUNCTION: STORAGE PROTEIN WHICH PROVIDES NITROGEN AND CARBON  
 CC RESERVES DURING GERMINATION AND SEEDLING GROWTH.  
 CC -1- MASS SPECTROMETRY: MW=56693.39; METHOD=MALDI.  
 CC -1- SIMILARITY: SOME, TO 75 SEED STORAGE PROTEINS.  
 KM Seed storage protein.  
 FT DISULFID 12 33  
 FT DISULFID 16 29  
 SO SEQUENCE 47 AA; 5698 MW; 588B0EC82273AC05 CRC64;

Query Match 33.3%; Score 86; DB 1; Length 47;  
 Best Local Similarity 34.1%; Pred. No. 0.024;  
 Matches 14; Conservative 11; Mismatches 16; Indels 0; Gaps 0;

QY 5 PROOYEQCCRCSEATEEREOCCRCRCREYKEDQROOE 45  
 DB 5 PRTEYACRVACQVAHEGVERORRCQOVCEKRLREGRRE 45

RESULT 5  
 TRHY\_HUMAN STANDARD; PRT; 1898 AA.  
 AC 007283;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE TRICHOHYALIN.  
 GN THH OR TRHY OR THL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93280194; PubMed=7685034;  
 RA Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D.,  
 Steinert P.M.;  
 RT "The structure of human trichohyalin. Potential multiple roles as a  
 functional EF-hand-like calcium-binding protein, a cornified cell  
 envelope precursor, and an intermediate filament-associated (cross-  
 linking) protein.";  
 RT J. Biol. Chem. 268:12164-12176(1993).  
 RN [2]  
 RP SEQUENCE OF 1731-1898 FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=93315897; PubMed=7686953;  
 RA O'Keefe E.J., Hamilton E.H., Lee S.-C., Steinert P.M.;  
 RT "Trichohyalin: a structural protein of hair, tongue, nail, and  
 epidermis";  
 RT J. Invest. Dermatol. 101:658-715(1993).  
 CC -1- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES  
 CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE  
 CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR  
 CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY  
 CC ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER  
 CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN  
 CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN  
 CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL  
 CC DIFFERENTIATION.  
 CC -1- SUBUNIT: MONOMER (PROBABLE).  
 CC -1- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS  
 CC THE INNER ROOT SHEAT (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN  
 CC THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF  
 CC THE EPIDERMIS.  
 CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND  
 CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST  
 CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS  
 CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED  
 CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS  
 CC THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.  
 CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN

CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG  
 CC DIFFERENT SPECIES.  
 CC -1- P.TM: KNOWN SUBSTATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE  
 CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMINASE.  
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE S-100  
 CC FAMILY.  
 CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: I09190; AAA65582.1; -  
 CC PIR: A45973; A45973.  
 CC HSSP: P02633; 1B0C.  
 CC MIM: 190370; -  
 CC DR INTERPRO: IPR001751; -  
 CC DR INTERPRO: IPR002048; -  
 CC DR PRAM: PR01023; S\_100; 1.  
 CC DR PFAM: PF00036; efhand; 1.  
 CC DR PROSITE: PS00018; EF\_HAND; 1.  
 CC DR PROSITE: PS00303; S100\_CABP; 1.  
 CC KM Repeat: Calcium-binding.  
 CC FT DOMAIN 1 91  
 CC FT CA\_BIND 22 33  
 CC FT CA\_BIND 62 73  
 CC FT DOMAIN 314 390  
 CC FT REPEAT 314 326  
 CC FT REPEAT 327 339  
 CC FT REPEAT 340 351  
 CC FT REPEAT 352 364  
 CC FT REPEAT 365 377  
 CC FT REPEAT 378 390  
 CC FT DOMAIN 391 444  
 CC FT REPEAT 391 396  
 CC FT REPEAT 397 402  
 CC FT REPEAT 403 408  
 CC FT REPEAT 409 414  
 CC FT REPEAT 415 420  
 CC FT REPEAT 421 426  
 CC FT REPEAT 427 432  
 CC FT REPEAT 433 438  
 CC FT REPEAT 439 444  
 CC FT DOMAIN 444 702  
 CC FT REPEAT 923 952  
 CC FT REPEAT 953 982  
 CC FT REPEAT 983 1012  
 CC FT REPEAT 1013 1042  
 CC FT REPEAT 1043 1072  
 CC FT REPEAT 1073 1102  
 CC FT REPEAT 1103 1132  
 CC FT REPEAT 1133 1162  
 CC FT DOMAIN 1250 1849  
 CC FT CONFLICT 1752 1752  
 CC FT CONFLICT 1794 1801  
 CC FT CONFLICT 1857 1857  
 CC FT CONFLICT 1880 1880  
 CC SO SEQUENCE 1898 AA; 247219 MW; A74B5947FB62E31D CRC64;

Query Match 33.3%; Score 86; DB 1; Length 1898;  
 Best Local Similarity 40.9%; Pred. No. 0.59;  
 Matches 18; Conservative 9; Mismatches 17; Indels 0; Gaps 0;

QY 2 ERDPROOYEQCCRCSEATEEREOCCRCRCREYKEDQROOE 45  
 DB 660 ERNNEQLRRROOERRRORLKRREBERERLORLKRREBERERROE 703

```

RESULT 6
ID TRHY_RABIT STANDARD: PRT: 1407 AA.
AC P37709:
DT 01-OCT-1994 (rel. 30, Created)
DT 01-OCT-1994 (rel. 30, Last sequence update)
DT 01-OCT-2000 (rel. 40, Last annotation update)
DE TRICHOHYALIN.
GN THH.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RA Fietz M.J., Rogers G.E.;
RU Submitted (DEC-1992) to the EMBL/Genbank/DBJ databases.
CC -I- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
CC ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
CC DIFFERENTIATION.
CC -I- SUBUNIT: HOMODIMER (PROBABLE).
CC -I- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
CC THE INNER ROOT SHEAT (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
CC THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
CC -I- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
CC THE EPIDERMIS.
CC -I- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED
CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
CC THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
CC DIFFERENT SPECIES.
CC -I- PTM: KNOWN SUBSTRATE OF TRANSGUTAMINASE. SOME 200 ARGININES ARE
CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMINASE.
CC -I- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100
CC FAMILY.
CC -I- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Z19092; CAA79519.1; -
DR PIR: S28589; S28589.
DR HSSP: P02633; 1BOC.
DR INTERPRO: IPR001751; -
DR INTERPRO: IPR002048; -
DR PFAM: PF01023; S.100; 1.
DR PFAM: PF00036; ehand; 1.
DR PROSITE: PS00018; EF_HAND; 1.
DR PROSITE: PS00303; S100_CABP; 1.
KW Repeat: Calcium-binding.
FT CA_BIND 22 33 SITE I (LOW AFFINITY) (POTENTIAL).
FT CA_BIND 62 73 SITE II (HIGH AFFINITY) (POTENTIAL).
SO SEQUENCE 1407 AA; 183781 MW; AE17D2A159F12B7F CRC64;
Query Match 32.8%; Score 84.5; DB 1; Length 1407;

```

```

Best Local Similarity 44.4%; Pred. No. 0.62;
Matches 24; Conservative 10; Mismatches 11; Indels 9; Gaps 4;
QY 2 ERDPR---QOY--ECCORCESEATEEREEOCCGEORCEKREKQEO--RQOEE 47
DB 1304 ERDRIAEQFARERKRELRKQEEQRRRRER-ERKFEQRLRQOEE 1356
RESULT 7
ID APA4_PAPAN STANDARD: PRT: 401 AA.
AC Q28758:
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 15-DEC-1998 (rel. 37, Last annotation update)
DE APOLIPOPROTEIN A-IV PRECURSOR (APO-AIV) (FRAGMENT).
GN APOA4.
OS Papio anubis (Olive baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Papio.
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=INTESTINE;
RX MEDLINE=93340170; PubMed=8101842;
RA Hixson J.E., Kammerer C.M., Molt G.E., Britten M.L., Birnbaum S.,
Powers P.K., Vandeberg J.L.;
RT "Baboon apolipoprotein A-IV. Identification of Lys-76-->Glu that
RT distinguishes two common isoforms and detection of length
RT polymorphisms at the carboxyl terminus."
CC -I- FUNCTION: MAY HAVE A ROLE IN CHYLOMICRONS AND VLDL SECRETION AND
CC CATABOLISM. REQUIRED FOR EFFICIENT ACTIVATION OF LIPOPROTEIN
CC LIPIASE BY APOC-II. POTENT ACTIVATOR OF LCAT. APOA-IV IS A MAJOR
CC COMPONENT OF HDL AND CHYLOMICRONS.
CC -I- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -I- TISSUE SPECIFICITY: SYNTHESIZED PRIMARILY IN THE INTESTINE.
CC SECRETED IN PLASMA.
CC -I- DOMAIN: NINE OF THE THIRTEEN 22-AMINO ACID TANDEM REPEATS (EACH
CC 22-MER IS ACTUALLY A TANDEM ARRAY OF TWO, A AND B, RELATED 11-
CC MERS) OCCURRING IN THIS SEQUENCE ARE PREDICTED TO BE HIGHLY ALPHA-
CC HELICAL, AND MANY OF THESE HELICES ARE AMPHIPATHIC. THEY MAY
CC THEREFORE SERVE AS LIPID-BINDING DOMAINS WITH LECITHIN:CHOLESTEROL
CC ACYLTRANSFERASE (LCAT) ACTIVATING ABILITIES.
CC -I- POLYMORPHISM: THERE ARE TWO COMMON APOA-IV ISOFORMS, I (SHOWN
CC HERE) AND E. THE I ISOFORM IS ASSOCIATED WITH HIGHER LEVELS OF
CC HIGH DENSITY LIPOPROTEIN-C ON A HIGH CHOLESTEROL, SATURATED FAT
CC DIET.
CC -I- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: L13174; AAA35379.1; -
DR HSSP: P02649; INFO.
DR INTERPRO: IPR000074; -
DR PFAM: PF01442; Apolipoprotein; 1.
KW Plasma; Lipid transport; HDL; Chylomicron; Repeat; Signal;
KW Polymorphism.
FT SIGNAL 1 1
FT CHAIN 5 401 POTENTIAL.
FT DOMAIN 17 314 APOLIPOPROTEIN A-IV.
FT REPEAT 17 38 13 X 22 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 44 65 2.
FT REPEAT 66 87 3.
FT REPEAT 99 120 4.
FT REPEAT 121 142 5.

```

```

FT REPEAT 143 164 6.
FT REPEAT 165 186 7.
FT REPEAT 187 208 8.
FT REPEAT 209 230 9.
FT REPEAT 231 252 10.
FT REPEAT 253 270 11.
FT REPEAT 271 292 12.
FT REPEAT 293 314 13.
FT DOMAIN 356 394 GLU/GLN-RICH.
FT VARIANT 80 80 K -> E (IN ISOFORM E).
SQ SEQUENCE 401 AA; 46538 MW; 0A76D1284AA9837F CRC64;

```

Query Match 30.8%; Score 79.5; DB 1; Length 401;  
 Best Local Similarity 36.4%; Pred. No. 0.59;  
 Matches 16; Conservative 13; Mismatches 8; Indels 7; Gaps 1;

```

OY 4 DPROVEQCQRCESEATEEREDQCRCERKEKEDQROOEE 47
Db 356 EPEQOQEQOQE-----GEQOQEQOQEQOQEQOQEQOQEQ 392

```

```

RESULT 8
ABRA_PLAFC STANDARD; PRT; 743 AA.
AC P22620;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE 101 KDA MALARIA ANTIGEN (P101) (ACIDIC BASIC REPEAT ANTIGEN).
GN ABRA.
OS Plasmodium falciparum (isolate Camp / Malaysia).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN [1]
RP MEDLINE=88298794; PubMed=3042768;
RX Weber J.L., Lyon J.A., Wolff R.H., Hall T., Lowell G.H., Chulay J.D.;
RT "Primary structure of a Plasmodium falciparum malaria antigen located
at the merozoite surface and within the parasitophorous vacuole.";
RL J. Biol. Chem. 263:11421-11425(1988).
CC -1- SUBCELLULAR LOCATION: AT THE MEROZOITE SURFACE AND WITHIN THE
PARASITOPHOUS VACUOLE.
CC -1- PTM: NOT GLYCOSYLATED (PROBABLE).

```

```

CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: J03903; AAA29462.1; -
DR DR PIR: A29232; A29232.
KW Antigen; Malaria; Repeat.
FT DOMAIN 226 273 8 X 6 AA REPEATS OF [VT]-N-D-[ED]-[ED]-D.
FT REPEAT 226 231
FT REPEAT 232 237
FT REPEAT 238 243
FT REPEAT 244 249
FT REPEAT 250 255
FT REPEAT 256 261
FT REPEAT 262 267
FT REPEAT 268 273
FT DOMAIN 674 731
SQ SEQUENCE 743 AA; 86622 MW; FBF0B0F8B07D922C CRC64;

```

Query Match 30.6%; Score 79; DB 1; Length 743;  
 Best Local Similarity 30.4%; Pred. No. 1.1;  
 Matches 14; Conservative 19; Mismatches 13; Indels 0; Gaps 0;

```

OY 2 ERDPROYEQCQRCESEATEEREDQCRCERKEKEDQROOEE 47

```

```

Db 678 EKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKE 723

```

```

RESULT 9
TAGB_DICDI STANDARD; PRT; 1905 AA.
AC P54683;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PRESTALK-SPECIFIC PROTEIN TAGB PRECURSOR (EC 3.4.21.-).
GN TAGB.
OS Dictyostelium discoideum (slime mold).
OC Eukaryota; Dictyostellida; Dictyostellum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RX MEDLINE=95262903; PubMed=7744252;
RA Shaulsky G., Kuspa A., Loomis W.F.;
RT "A multidrug resistance transporter/serine protease gene is required
for prestalk specialization in Dictyostelium.";
RL Gene Dev. 9:1111-1122(1995).
CC -1- FUNCTION: INTERCELLULAR COMMUNICATION VIA TAGB MAY MEDIATE
CC INTEGRATION OF CELLULAR DIFFERENTIATION WITH MORPHOGENESIS.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO PEPTIDASE FAMILY
CC S8; ALSO KNOWN AS THE SUBTILASE FAMILY.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ATP-BINDING
CC TRANSPORT PROTEIN FAMILY (ABC TRANSPORTERS). MDR SUBFAMILY.
CC -1- SIMILARITY: STRONG, TO TAGC.

```

```

CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U20433; AA62212.1; -
DR HSSP: P13569; INBD.
DR DICTYDB: DD02059; TAGB.
DR INTERPRO: IPR00209; -.
DR INTERPRO: IPR00140; -.
DR INTERPRO: IPR001617; -.
DR PPM: P00664; ABC membrane; 1.
DR PPM: P00005; ABC_tran; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; FALSE_NEG.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
KW Hydrolase; Serine protease; ATP-binding; Transport; Transmembrane;
Signal.
FT SIGNAL 1 7
FT CHAIN 1 1905
FT DOMAIN 378 700
FT DOMAIN 7 1905
FT TRANSMEM 1011 1031
FT TRANSMEM 1076 1096
FT TRANSMEM 1121 1141
FT TRANSMEM 1210 1230
FT TRANSMEM 1309 1329
FT TRANSMEM 1332 1352
FT ACT_SITE 387 387
FT ACT_SITE 432 432
FT ACT_SITE 695 695
FT ACT_SITE 1553 1560
FT NP_BIND 63 67
FT DOMAIN 95 104
FT DOMAIN 107 134
FT DOMAIN 311 321
FT DOMAIN 833 839

```

```

FT SIGNAL 1 7
FT CHAIN 1 1905
FT DOMAIN 378 700
FT DOMAIN 7 1905
FT TRANSMEM 1011 1031
FT TRANSMEM 1076 1096
FT TRANSMEM 1121 1141
FT TRANSMEM 1210 1230
FT TRANSMEM 1309 1329
FT TRANSMEM 1332 1352
FT ACT_SITE 387 387
FT ACT_SITE 432 432
FT ACT_SITE 695 695
FT ACT_SITE 1553 1560
FT NP_BIND 63 67
FT DOMAIN 95 104
FT DOMAIN 107 134
FT DOMAIN 311 321
FT DOMAIN 833 839

```

FT	DOMAIN	838	844	POLY-GLY.
FT	DOMAIN	876	876	POLY-LEU.
FT	DOMAIN	1012	1015	POLY-TLE.
FT	DOMAIN	1386	1389	POLY-GLU.
FT	DOMAIN	1398	1404	POLY-GLY.
FT	DOMAIN	1445	1450	POLY-ASN.
FT	DOMAIN	1765	1779	POLY-ASN.
FT	DOMAIN	1782	1785	POLY-SER.
FT	DOMAIN	1807	1812	POLY-PRO.
FT	DOMAIN	1815	1860	POLY-GLN.
FT	DOMAIN	1872	1878	POLY-PRO.
FT	CARBOHYD	594	594	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	621	621	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	672	672	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	747	747	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	823	823	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	1172	1172	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	1522	1522	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	1658	1658	N-LINKED (GLCNAC. . .) (POTENTIAL)
50	SEQUENCE	1905 AA:	212518 MW:	BBE223FA8B9AE13C CRC64;

Query Match	30.68;	Score 79;	DB 1;	Length 1905;
Best Local Similarity	25.68;	Pred. No. 2.5;		
Matches 11; Conservative	23;	Mismatches 9;	Indels 0;	Gaps 0

```
QY      5 PROXYEQCORCESEATEEREQEQCQRRCERKEYKEQQRQEE 47
          |::|||::: : ::::|:|:|: : ::::|:|:|:
Db     1812 PQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQ 1854
```

RESULT	10
IF2_SALTY	
ID	IF2_SALTY
STANDARD;	
PRT;	892 AA

DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE TRANSLATION INITIATION FACTOR IF-2.  
 GN *INFB*.  
 OS *Salmonella typhimurium*.  
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

CC salmonella.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-172;  
 RA Page-larsen J., Steffensen S.A.D.A., Hedegaard J., Olsen J.E.,  
 RA Mortensen K.K., Speerling-Petersen H.U.;  
 RT "Sequence of the infB gene from Salmonella typhimurium";  
 RL submitted (NOV-1997) to the EMBL/Genbank/DBD databases.  
 CC -1- FUNCTION: IF-2, ONE OF THE ESSENTIAL COMPONENTS FOR THE INITIATION  
 CC OF PROTEIN SYNTHESIS IN VITRO, PROTECTS FORMYLMETHIONYL-tRNA FROM  
 CC SPONTANEOUS HYDROLYSIS AND PROMOTES ITS BINDING TO THE 30S  
 CC RIBOSOMAL SUBUNIT. IT IS ALSO INVOLVED IN THE HYDROLYSIS OF GTP  
 CC DURING THE FORMATION OF THE 70S RIBOSOMAL COMPLEX (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- ALTERNATIVE PRODUCTS: USING ALTERNATIVE INITIATION CODONS IN THE  
 CC SAME READING FRAME, THE GENE TRANSLATES INTO THREE ISOZYMES:  
 CC ALPHA, BETA AND GAMMA.  
 CC -1- SIMILARITY: BELONGS TO THE IF-2 FAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/isb-ch>).  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC -----

CC -----  
DR EMBL: AJ002552; CAA05549.1; -  
DR EMBL: AJ002552; CAA05550.1; -  
DR EMBL: AJ002552; CAA05551.1; -  
DR STYGENE: SGI0757; INF.

DR	INTERPRO: IPR000178; -			
DR	INTERPRO: IPR000795; -			
DR	PFAM: PF00009; GTP_EFTU; 1.			
DR	PFAM: PF02131; IF2; 1.			
DR	PROSITE: PS01176; IF2; 1.			
KW	Initiation factor; protein biosynthesis; GTP-binding; Alternative initiation.			
KW	Alternative initiation.			
FT	CHAIN	1	892	
FT	CHAIN	159	892	INITIATION FACTOR IF-2-ALPHA
FT	CHAIN	167	892	INITIATION FACTOR IF-2-BETA.
FT	INIT_MET	159	139	INITIATION FACTOR IF-2-GAMMA
FT	INIT_MET	167	157	FOR IF2-BETA.
FT	DOMAIN	167	167	FOR IF2-GAMMA.
FT	DOMAIN	394	542	G-DOMAIN.
FT	NP_BIND	400	407	GTP (BY SIMILARITY).
FT	NP_BIND	446	503	GTP (BY SIMILARITY).
FT	NP_BIND	500	503	GTP (BY SIMILARITY).
SO	SEQUENCE	892 AA;	97388 MW;	9472B20DD55597BI CRC64;

Query Match	30.2%;	Score 78;	DB 1;	Length 892;
Best Local Similarity	33.3%;	Pred. No. 1.6;		
Matches	18;	Conservative	12;	Mismatches 16;
				Indels 8;
				Gaps 2.

```
OY      2 ERPRQ-----QECCQRCESEATEBEREQ---QCERCFEYKEGGRQEEE    47
```

::: | | | : | :: : : | | | :: :

```
Db     94 KDPQAERLAEEEAQQAEAAEEGARREAAEQAKAEEAAEQARKRAAE   147
```

```

RESULT 11
SNF5_YEAST
ID SNF5_YEAST STANDARD; PRT; 905 AA

```

AC P1480; (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 30, Last sequence update)  
 DT 01-OCT-1994 (Rel. 40, Last annotation update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE TRANSCRIPTION REGULATORY PROTEIN SNF5 (SNI/SNF COMPLEX COMPONENT SNF5  
 DE (TRANSCRIPTION FACTOR TYE4)).  
 GN SNF5 OR TYE4 OR SMI10 OR YBR289W OR YBR2036.  
 OS *Saccharomyces cerevisiae* (Baker's Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;  
 OC Saccharomycetaceae; Saccharomycetes.

RN SEQUENCE FROM N.A.  
 RI 111  
 RX STRAIN=MCV;  
 RC MEDLINE=91042489; Pubmed=2233708;  
 RA Laurent B.C., Treitel M.A., Carlson M.;  
 RT "rme SNF5 protein of Saccharomyces cerevisiae is a glutamine- and  
 RT proline-rich transcriptional activator that affects expression of a  
 RL Mol. Cell. Biol. 10:3616-5625(1990).

RN SEQUENCE FROM N.A.  
 RP  
 RC STRAIN-S288C;  
 RX MEDLINE=94378722; PubMed=8091861;  
 RA Holmstrom K., Brandt T., Killesoe T.;  
 RT "The sequence of a 32,420 bp segment located on the right arm of  
 RT chromosome II from *Saccharomyces cerevisiae*.";  
 RT Yeast 10:S47-S62(1994).

CC -1 FUNCTION: INVOLVED IN TRANSCRIPTIONAL ACTIVATION. THE SWI/SNF COMPLEX IS REQUIRED FOR THE INDUCED EXPRESSION OF A LARGE NUMBER OF GENES. THIS COMPLEX ALTERS CHROMATIN STRUCTURE TO FACILITATE BINDING OF GENE-SPECIFIC DEDICATED TRANSCRIPTION FACTORS.

CC -1 SUBUNIT: COMPONENT OF THE SWI/SNF GLOBAL TRANSCRIPTION ACTIVATOR COMPLEX.

CC -1 SUBCELLULAR LOCATION: NUCLEAR.

CC -1 SIMILARITY: BELONGS TO THE SNF5 FAMILY.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

[illegible]

CC		PRECURSORS IN THE CENTRAL AND PERIPHERAL NERVOUS SYSTEMS, BUT NOT FOR THE SPECIFICATION OF IDENTITY. PROSPERO PROTEIN MAY REGULATE TRANSCRIPTION BY BINDING TO DNA.
CC	-I-	SUBCELLULAR LOCATION: NUCLEAR.
CC	-I-	ALTERNATIVE PRODUCTS: 2 ISOFORMS; PROS-L (SHOWN HERE) AND PROS-S; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY HAVE DIFFERENT N-TERMINAL AMINO ACIDS OF THE HOMEODOMAIN.
CC	-I-	TISSUE SPECIFICITY: NEURONAL PRECURSORS. EXPRESSED IN THE DEVELOPING CNS, LENS-SECRETING CONE CELLS OF THE EYE, AND MIDGUT.
CC	-I-	DEVELOPMENTAL STAGE: EXPRESSED IN NEURONAL PRECURSORS EARLY DURING FORMATION.
CC	-I-	SIMILARITY: BELONGS TO THE PROSPERO FAMILY OF HOMEOBOX PROTEINS.
CC		-----
CC		This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on ways CC can be used by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch">http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch</a> ).
CC		-----
DR	EMBL;	M81389; AAA28841.1; ;
DR	EMBL;	D10609; BAA01464.1; ;
DR	EMBL;	Z11743; CAA77802.1; ;
DR	PIR;	A41089; A41089.
DR	PIR;	JQ1397; JQ1397.
DR	FlyBase;	FBgn0004595; pros.
KW	Nuclear protein;	Transcription regulation; DNA-binding; Homeobox; developmental protein; Alternative splicing.
FT	DOMAIN	4 12
FT	DOMAIN	28 31
FT	DOMAIN	32 35
FT	DOMAIN	188 191
FT	DOMAIN	253 260
FT	DOMAIN	270 276
FT	DOMAIN	282 286
FT	DOMAIN	431 437
FT	DOMAIN	505 508
FT	DOMAIN	717 737
FT	DOMAIN	754 763
FT	DOMAIN	766 772
FT	DOMAIN	934 937
FT	DOMAIN	952 957
FT	DOMAIN	960 963
FT	DOMAIN	966 970
FT	DOMAIN	991 998
FT	DOMAIN	1029 1048
FT	DOMAIN	1074 1082
FT	DNA_BIND	1241 1303
FT	DOMAIN	1316 1403
FT	VASSP_LIC	1204 1244
FT	CONFLICT	76 98
FT		
FT	CONFLICT	120 144
FT		
FT	CONFLICT	418 418
FT	CONFLICT	677 677
FT	CONFLICT	802 802
FT	CONFLICT	958 958
FT	CONFLICT	1048 1048
SO	SEQUENCE	1403 AA: 153569 MW: 9EBD973E24E238E CRC64:

```
Query Match          30.2%; Score 78; DB 1; Length 1403;
Best Local Similarity 29.4%; Pred No. 2.4;
Matches    15; Conservative   19; Mismatches   13; Indels    4; Gaps     1

QY      1 YERDFRQOYEQCQRRCESATEEREDEDCEQ----RCIEREYKQQROQEEH 47
       |:| | | : : : : |:| | | | |||||:|:|
Db      704 FGEGRMAKKEALEEQDQDQDQDQDQDQDQDQDQDQRRFFEDQEGDQRRKEEO 754
```

## RESULT 13

	SKSL_YEAST	STANDARD:	PRT:	502 AA.
AC	Q12505.7			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	SERINE/THEONINE-PROTEIN KINASE SKS1 (EC 2.7.1.-) (SUPPRESSOR KINASE OF SNF3).			
GN	SKS1 OR SHA3 OR YPIL026C OR LPP5.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;			
CC	Saccharomycetaceae; Saccharomyces.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=97103770; PubMed=8948096;			
RA	Tang Z., Blisson L.F.;			
RT	"The SKS1 protein kinase is a multicopy suppressor of the snf3 mutation of Saccharomyces cerevisiae.";			
RL	Yeast 12:1407-1419(1996).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Wang Y., Ahmed A., Bussey H., Fortin N., Friesen J.D., Hall J., Storms R.K., Vo D.H., Winnett E.;			
RL	Submitted (SEP-1995) to the EMBL/Genbank/DBJ databases.			
CC	-I- FUNCTION: MAY HAVE A ROLE IN GLUCOSE REGULATION			
CC	-I- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-PROTEIN KINASES.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - its the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; U30613; AAC49570.1; -			
DR	HSSP; P24941; IAO1.			
DR	SGD; S0005947; SKS1.			
DR	INTERPRO; IPR000719; -			
DR	INTERPRO; IPR002290; -			
DR	PFAM; PF00069; Pkinase; 1.			
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP_1.			
DR	PROSITE; PS00108; PROTEIN_KINASE_ST_1.			
DR	PROSITE; PSS0011; PROTEIN_KINASE_DOM_1.			
KW	Transferrase; Serine/threonine-protein kinase; ATP-binding.			
FT	DOMAIN 10 338 PROTEIN KINASE			
FT	NP BIND 16 24 ATP (BY SIMILARITY).			
FT	BINDING 39 39 ATP.			
FT	ACT_SITE 186 186 BY SIMILARITY.			
FT	MUTAGEN 39 39 K->R: LOSS OF ACTIVITY.			
SO	SEQUENCE 502 AA; 57844 MW; B106D084BAFA61E5 CRC64;			
OY	1 YERPPOQYEECCRCSEATERREHQCEBCRCHEHYEQRRQEE 47			
Db	372 FSADQEQDQQDQQD--DQQDYQEQEQEKQKQIQIONQEQADQDQDEED 415			
	Query Match 30.0%; Score 77.5; DB 1; Length 502;			
	Best Local Similarity 31.9%; Pred. No. 1.1;			
	Matches 15; Conservative 15; Mismatches 14; Indels 3; Gaps 1;			
	RESULT 14			
ID	TF2D_HUMAN	STANDARD;	PRT;	339 AA.
AC	P20226;			
DT	01-FEB-1991 (Rel. 17, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	15-JUL-1998 (Rel. 36, Last annotation update)			
DE	TRANSCRIPTION INITIATION FACTOR TFIID (TATA-BOX FACTOR) (TATA SEQUENCE-BINDING PROTEIN) (TBP).			

TNB OR TFIIID OR TF2D.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 MN Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90302010; Pubmed=2194289;  
 RA Kao C.C., Lieberman P.M., Schmidt M.C., Zhou Q., Pei R., Berk A.J.;  
 RL "Cloning of a transcriptionally active human TATA binding factor."; *Science* 248:1646-1649(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90326195; Pubmed=2374612;  
 RA Hoffmann A., Sinn E., Yamamoto T., Wang J., Roy A., Horikoshi M.,  
 RA Roeder R.G.;  
 RL "Highly conserved core domain and unique N terminus with presumptive  
 regulatory motifs in a human TATA factor (TFIID)."; *Nature* 346:387-390(1990).  
 RN [3]  
 RP DOMAINS.  
 RX MEDLINE=90302006; Pubmed=2363050;  
 RA Peterson M.G., Tanese N., Pugh B.F., Tjian R.;  
 RL "Functional domains and upstream activation properties of cloned  
 human TATA binding protein."; *Science* 248:1625-1630(1990).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 159-337.  
 RX MEDLINE=96209823; Pubmed=8643494;  
 RA Nikolov D.B., Chen H., Halay E.D., Hoffmann A., Roeder R.G.,  
 RA Burley S.K.;  
 RL "Crystal structure of a human TATA box-binding protein/TATA element  
 complex."; *Proc. Natl. Acad. Sci. U.S.A.* 93:4862-4867(1996).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 159-339.  
 RX MEDLINE=96346176; Pubmed=8757291;  
 RA Joo Z.S., Chiu T.K., Leiberman P.M., Baikalov I., Berk A.J.,  
 RA Dickerson R.E.;  
 RL "How proteins recognize the TATA box."; *J. Mol. Biol.* 261:239-254(1996).  
 CC -1- FUNCTION: GENERAL FACTOR THAT PLAYS A MAJOR ROLE IN THE ACTIVATION  
 OF EUKARYOTIC GENES TRANSCRIBED BY RNA POLYMERASE II. TFIIID BINDS  
 SPECIFICALLY TO THE TATA BOX PROMOTER ELEMENT WHICH LIES CLOSE TO  
 THE POSITION OF TRANSCRIPTION INITIATION.  
 CC -1- SUBUNIT: BINDS DNA AS A MONOMER.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: THE C-TERMINAL 180 RESIDUES ARE EXTREMELY WELL  
 CONSERVED IN ALL EUKARYOTIC TFIIID.  
 CC -1- SIMILARITY: WEAK, WITH BACTERIAL POLYMERASE SIGMA-FACTORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M556953; AAA36731.1; -;  
 DR EMBL; X549633; CA38736.1; -;  
 DR PIR; A34830; A34830.  
 DR PIR; A34831; A34831.  
 DR PDB; 1TGH; 01-AUG-96.  
 DR PDB; 1CDW; 23-DEC-96.  
 DR TRANSFAC; T00794; -;  
 DR MIM; 600075; -;  
 DR INTERPRO; IPR000814; -;  
 DR PFAM; PF00352; TBP; 2.  
 DR PRINTS; PRO0686; TIFACTORIID.  
 DR PROSITE; PS00351; TFIIID; 2.  
 KW Transcription regulation; DNA-binding; Nuclear protein; Repeat;  
 TM 3D-structure.



```

FT REPEAT      165      241      1.
FT REPEAT      255      332      2.
FT CONFLICT    58       61      MISSING (IN REF. 2).
SO SEQUENCE    339 AA; 37698 MW; A61A578D972B970B CRC64;

Query Match
Best Local Similarity 30.2%; Score 77; DB 1; Length 339;
Matches 13; Conservative 18; Mismatches 12; Indels 0; Gaps 0;

QY 2 ERDPQOYEQRCRCESEATEER--EQEQCEQRCEREXKQEQQEQ 44
   | 111 : 1 : : : : : : : : : : : : : : : : : : : :
Db 53 EEQGRQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 95

RESULT 15
NFM_CHICK
ID NFM_CHICK STANDARD; PRT; 857 AA.
AC P16053;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEUROFILAMENT TRIPLET M PROTEIN (160 KDA NEUROFILAMENT PROTEIN)
DE (NEUROFILAMENT MEDIUM POLYPEPTIDE) (NF-M).
GN NFM.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90174973; PubMed=2106668;
RA Zopf D., Dineva B., Betz H., Gundelfinger E.D., Betz H.;
RT "Isolation of the chicken middle-molecular weight neurofilament
   (NF-M) gene and characterization of its promoter.";
RT Nucleic Acids Res. 18:521-529(1990).
RN [2]
RP SEQUENCE OF 259-857 FROM N.A.
RX MEDLINE=88112814; PubMed=3123320;
RA Zopf D., Hermans-Borgmeyer I., Gundelfinger E.D., Betz H.;
RT "Identification of gene products expressed in the developing chick
   visual system: characterization of a middle-molecular-weight
   neurofilament cDNA.";
RL Genes Dev. 1:699-708(1987).
CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFM IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X17102; CAA34958.1; -
DR EMBL: X05558; CAA29073.1; -
DR PIR: A27040; A27040.
DR PIR: S08061; S08061.
DR PIR: S15762; S15762.
DR INTERPRO: IPR001664; -
DR PRAM: PR00038; filament; 1.
DR PROSITE: PS00226; IF; 1.

```

```

KW Intermediate filament; Heptad repeat pattern; Coiled coil; Neurone;
KW Phosphorylation; Glycoprotein.
FT INIT MET 0
FT DOMAIN 1 98 HEAD.
FT DOMAIN 99 406 ROD.
FT DOMAIN 407 857 TAIL.
FT DOMAIN 99 130 COIL 1A.
FT DOMAIN 131 143 LINKER 1.
FT DOMAIN 144 242 COIL 1B.
FT DOMAIN 243 259 LINKER 12.
FT DOMAIN 260 281 COIL 2A.
FT DOMAIN 282 285 COIL 2B.
FT DOMAIN 286 406 COIL 2B.
FT CARBOHYD 46 46 O-LINKED (GLCNAC) (BY SIMILARITY).
FT CARBOHYD 426 426 O-LINKED (GLCNAC) (BY SIMILARITY).
FT CONFLICT 546 546 G -> R (IN REF. 2).
SO SEQUENCE 857 AA; 95704 MW; 4E2F0FCAC64778B CRC64;

Query Match
Best Local Similarity 39.5%; Score 77; DB 1; Length 857;
Matches 17; Conservative 8; Mismatches 16; Indels 2; Gaps 1;

QY 7 QQYEQRCRCESEATEER--EQEQCEQRCEREXKQEQQEQ 47
   | 11 : 1 : 1 1 1 : : 1 1 : : : : : : : : : :
Db 481 QEEQEEKAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 523

```

Search completed: March 1, 2001, 16:16:52  
Job time: 432 sec

